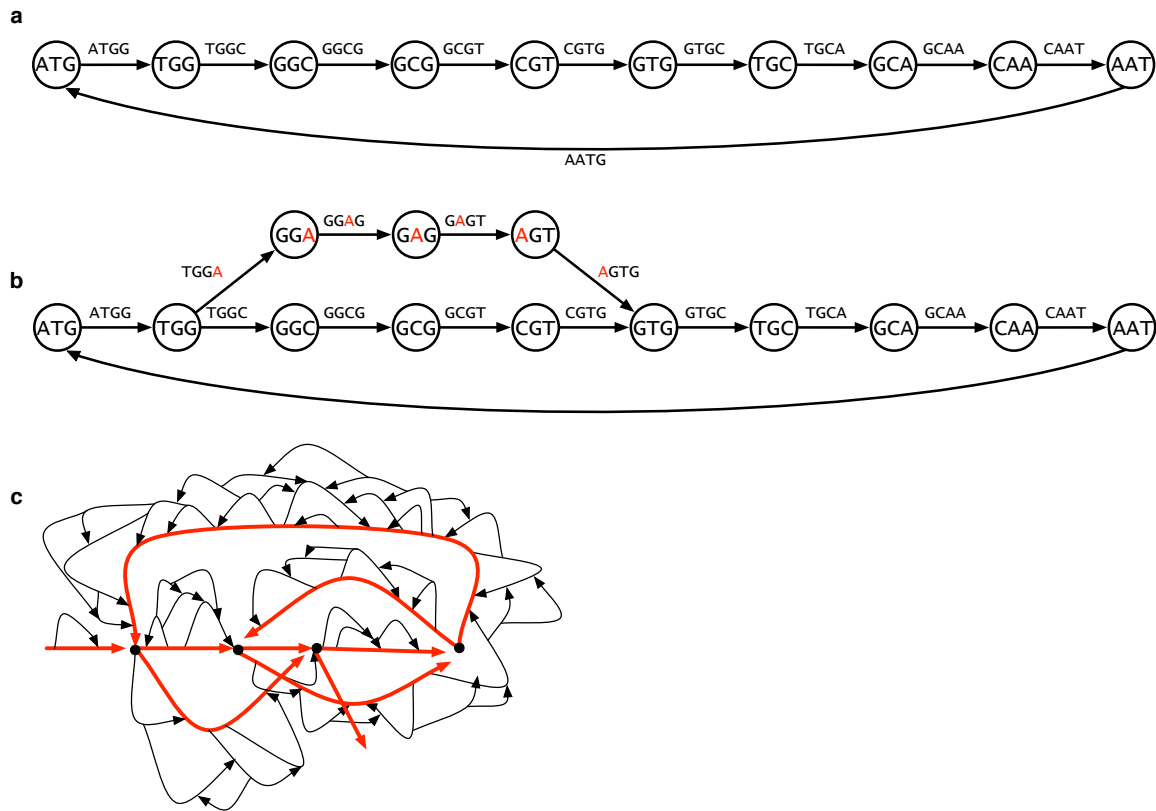


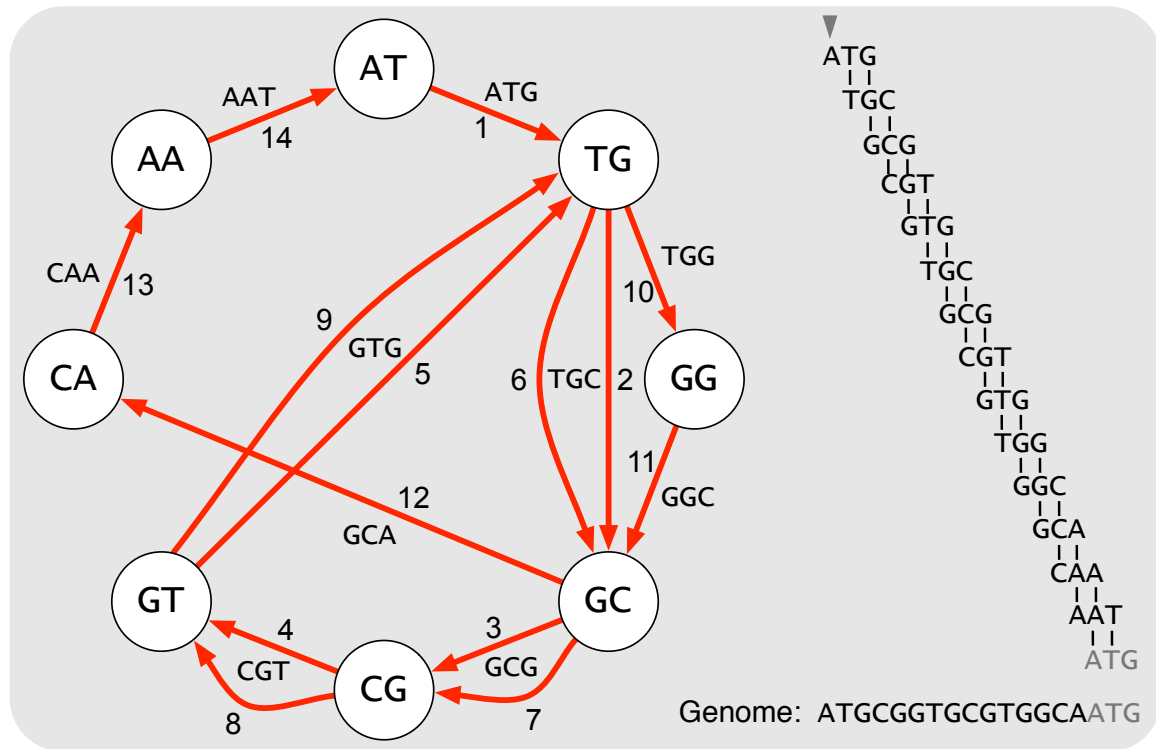
Supplementary Figures

Why are de Bruijn graphs useful for genome assembly?

Phillip E. C. Compeau, Pavel A. Pevzner & Glenn Tesler



Supplementary Figure 1. De Bruijn graph from reads with sequencing errors. (a) A de Bruijn graph E on our set of reads with $k = 4$. Finding an Eulerian cycle is already a straightforward task, but for this value of k , it is trivial. (b) If TGGAGTG is incorrectly sequenced as a sixth read (in addition to the correct TGGCGTG read), then the result is a *bulge* in the de Bruijn graph, which complicates assembly. (c) An illustration of a de Bruijn graph E with many bulges. The process of bulge removal should leave only the red edges remaining, yielding an Eulerian path in the resulting graph.



Supplementary Figure 2. De Bruijn graph of a genome with repeats. The graph E for k -mers with different multiplicities: each of the four 3-mers TGC, GCG, CGT, and GTG has multiplicity 2, and each of the six 3-mers ATG, TGG, GGC, GCA, CAA, and AAT has multiplicity 1. An Eulerian cycle is formed by following the numbered edges in the order 1,2,...,14: **ATG**, **TGC**, **GCG**, **CGT**, **GTG**, **TGC**, **GCG**, **CGT**, **GTG**, **TGG**, **GGC**, **GCA**, **CAA**, **AAT**. This Eulerian cycle spells the cyclic superstring **ATGCGGTGCGTGGCA**.